

INSERM, Biostatistique

at Bordeaux,

FRANCE

Review: Evaluating time to cancer recurrence as
a surrogate marker for survival from an
information theory perspective

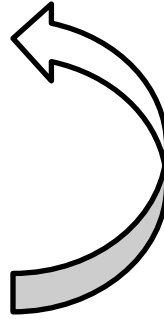
by Alonso and Molenberghs,

Statistical Methods in Medical Research

2008; **17**, 497-504.

Choice of endpoints

- T = True endpoint
(difficult to measure)
costly; long duration time
- S = Surrogate endpoint
(easy to measure)



Replace?

→ High association

Example: Colon cancer meta-analysis

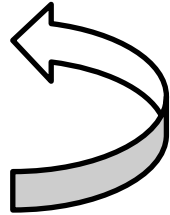
(Sargent et al. 2005)

- T = Overall survival (death due to any cause)
- S = Time to cancer recurrence (easy to measure)
Surrogate to assess the treatment effect on T.

- Linear regression

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}}$$



Predict well?

→ High correlation

- Coefficient of determination

$$R^2 = \frac{\|\mathbf{y} - \bar{y}\|^2 - \|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\|\mathbf{y} - \bar{y}\|^2}$$

$$= \text{corr}^2(\mathbf{y}, \hat{\mathbf{y}})$$

- : $R^2 \approx 1$

$$\Leftrightarrow \text{corr}^2(\mathbf{y}, \hat{\mathbf{y}}) \approx 1$$

$\Leftrightarrow \hat{\mathbf{y}}$ is a good predictor of \mathbf{y}

In practice, e.g., $R^2 > 0.8$

Criteria of good surrogate

- Treatment indicator

$$Z = \begin{cases} 0 & \text{for control} \\ 1 & \text{for treatment} \end{cases}$$

- Treatment effect on the true endpoint

$$T | Z$$

- Treatment effect on the surrogate endpoint

$$S | Z$$

- **Freedman et al. (1992)'s scheme**

S is a good surrogate of T

if $(T | Z)$ is explained by $(S | Z)$

→ Statistical validation is difficult
without aid of **meta-analysis**

Replace?



Meta-analytic assessment of surrogate

- Use meta-analysis or multi-center trials for validating a surrogate

Daniels et al. (1997), Albert et al. (1998),
Gail et al. (2000), Buyse et al. (2000)

- Buyse et al. (2000) introduce a meta-analytic definition:

R_{Trial}^2 --- trial level
(relatively straightforward to calculate)

R_{Ind}^2 --- individual level
(several different ways to define

→ different setting yield different definitions)

- The present paper review an information-theoretic definition for R_{Ind}^2 as a unified system

Information theory

$X \sim \text{pdf } f(X)$

$h(X) = -E \log f(X)$: Entropy → Represent uncertainty

- Power Entropy $EP(X) = \frac{1}{2\pi e} e^{2h(X)}$

Example 1:

$$X \sim N(\mu, \sigma^2) \Rightarrow EP(X) = \sigma^2 = \text{Var}(X)$$

Example 2:

$$X \sim \text{pdf } f(X) = \frac{1}{\mu} e^{-\frac{X}{\mu}} \Rightarrow EP(X) = e^2 \mu^2 \propto \text{Var}(X)$$

Summary:

Power entropy represent **uncertainty** about X

Information theory

- The surrogate (S) is a good surrogate for the true endpoint (T) if uncertainty about T is largely reduced by S

$$EP(T | Z) \gg EP(T | Z, S)$$

Similar to

$$Var(X) > EVar(X | Y)$$

- S is useless surrogate for T

$$EP(T | Z) = EP(T | Z, S) \quad \text{or} \quad (T \perp S | Z)$$

- Information theoretic definition of R_squared

$$R_h^2 = \frac{EP(T | Z) - EP(T | Z, S)}{EP(T | Z)}$$

$$R_h^2 = 0 \Leftrightarrow S \perp T | Z,$$

$$R_h^2 = 1 \Leftrightarrow \text{No uncertainty about } T \text{ if we know } S$$

Case study

Example: Advanced colon cancer meta-analysis (10 trials)

- T = Overall survival (death due to any cause)
- S = Time to cancer recurrence (easy to measure)
- Z = Treatment (CP only vs. CP & CAP)

Validation of trial level surrogacy (Method I):

Step 1: Fit separate Cox regressions

$$\begin{cases} r_{ij}(t | Z_{ij}) = r_{0i}(t) \exp(\alpha_i Z_{ij}) & (\text{time-to-recurrence } S_{ij}) \\ \lambda_{ij}(t | Z_{ij}) = \lambda_{0i}(t) \exp(\beta_i Z_{ij}) & (\text{time-to-death } T_{ij}) \end{cases}$$

(trial-specific effects model on the treatment effect)

Step 2: $(\hat{\alpha}_i, \hat{\beta}_i), i = 1, \dots, 10$ are used to estimate $R_h^2 = R_{Trial}^2$

$$\text{Like } \hat{R}_{Trial}^2 = \text{corr}(\hat{\alpha}_., \hat{\beta}_.)^2 = 0.82$$

Case study

Estimation of trial level surrogacy (Method II):

Step 1: Fit shared frailty model

$$\begin{cases} r_{ij}(t | Z_{ij}) = u_i r_0(t) \exp(\alpha Z_{ij}) & (\text{time-to-recurrence } S_{ij}) \\ \lambda_{ij}(t | Z_{ij}) = u_i \lambda_0(t) \exp(\beta Z_{ij}) & (\text{time-to-death } T_{ij}) \end{cases}$$

(random effects model on the baselines)

Step 2: $(\hat{\alpha}_i = e^{u_i} \hat{\alpha}, \hat{\beta}_i = e^{u_i} \hat{\beta}), i = 1, \dots, 10$

are used to estimate $R_h^2 = R_{Trial}^2 = 0.88$

Estimation of individual level surrogacy

$$\begin{cases} r_{ij}(t | Z_{ij}) = r_0(t) \exp(\alpha_i Z_{ij}) & (\text{time-to-recurrence } S_{ij}) \\ \lambda_{ij}(t | Z_{ij}) = \lambda_0(t) \exp(\beta_{Si} Z_{ij} + \gamma_i I(S_{ij} \leq t)) & (\text{time-to-death } T_{ij}) \end{cases}$$

Using the method of [Alonso et al. \(2007\)](#), $R_{Ind}^2 = 0.76$ estimated

Summary

- Information theoretic R_h^2 is suggested,
 - interpretable as the dependence between S and T given Z.
 - estimable irrespective of data type
- In meta-analysis two types of R_h^2 exists.
 - 1) Trial level $R_h^2 = R_{Trial}^2$
 - ➔ Dependence between S_i and T_i
 - 2) Individual level $R_h^2 = R_{Ind}^2$
 - ➔ Dependence between S_ij and T_ij

My copula-based approach also try to incorporate
“individual-level” dependence via copulas

Joint frailty-copula model (Proposed)

Frailty model (Rondeau et al., 2011)

$$\begin{cases} r_{ij}(t | u_i) = u_i r_0(t) \exp(\boldsymbol{\beta}'_1 \mathbf{Z}_{ij}) & (\text{time - to - progression } X_{ij}) \\ \lambda_{ij}(t | u_i) = u_i^\alpha \lambda_0(t) \exp(\boldsymbol{\beta}'_2 \mathbf{Z}_{ij}) & (\text{time - to - death } D_{ij}) \end{cases}$$

+

Copula model :

$$\Pr(X_{ij} > x , D_{ij} > y | u_i) = C_\theta [\exp \{ - R_{ij}(x | u_i) \} , \exp \{ - \Lambda_{ij}(y | u_i) \}]$$

where C_θ is a copula (Nelsen, 2006), and

$$R_{ij}(x | u_i) = \int_0^x r_{ij}(v | u_i) dv ,$$

**Individual level
dependence**

$$\Lambda_{ij}(y | u_i) = \int_0^y \lambda_{ij}(v | u_i) dv$$